DOCUMENT RESUME

ED 171 783 95 TM 009 253

AUTHOR Schmidt, William H.
TITLE Measuring the Content of Instruction. Research Series No. 35.
INSTITUTION Michigan State Univ., East Lansing. Inst. for Research on Teaching.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C. Basic Skills Group. Teaching Div.
PUB DATE Oct 78
CONTRACT 400-76-0073
NOTE 19p.; Paper presented at the Annual Meeting of the American Educational Research Association (62nd, Toronto, Ontario, Canada, March 27-31, 1978); For related document, see ED 155 215
AVAILABLE FROM Institute for Research on Teaching, College of Education, Michigan State University, 252 Erickson Hall, East Lansing, Michigan 48824 ($1.50)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Achievement Tests; *Content Analysis; *Curriculum Evaluation; *Elementary School Mathematics; *Evaluation Criteria; Grade 4; Intermediate Grades; *Taxonomy; Test Validity; Textbook Evaluation
IDENTIFIERS California Test of Basic Skills; Iowa Tests of Basic Skills; Metropolitan Achievement Tests; Stanford Achievement Tests

ABSTRACT

A taxonomy was developed for classifying instructional materials in fourth grade mathematics, as well as the content of actual classroom instruction and of achievement tests. The taxonomy is defined by the intersection of three factors and results in 468 cells: (1) mode of presentation (for example, use of graphs, tables, story problems)--3 levels; (2) nature of the numbers (single digits, fractions, geometric figures, and so on)--13 levels; (3) arithmetic operations--12 levels. This system was used to classify the mathematics subtests of four widely-used standardized achievement tests: the Stanford Achievement Test (SAT); Iowa Test of Basic Skills (ITBS); Metropolitan Achievement Test (MAT); and California Test of Basic Skills (CTBS). Three raters classified the content of these four tests, with inter-rater reliability ranging from 60% to 90%. Results indicated sufficient variance across tests to suggest that some tests may have more content validity for a given instructional program than others. A total score on any mathematics test represents an aggregate of many content areas. Since the taxonomy was able to denote content differences among tests, it was concluded that it would also be useful for an analysis of instructional content. (GDC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

William H. Schmidt

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

ED171783

IMOU9 253

Institute for Research on Teaching

College of Education — Michigan State University

Research Series No. 35

MEASURING THE CONTENT OF INSTRUCTION

William H. Schmidt

Published By

The Institute for Research on Teaching
252 Erickson Hall
Michigan State University
East Lansing, Michigan 48824

October, 1978

The work reported herein is sponsored by the Institute for Research on Teaching, College of Education, Michigan State University. The Institute for Research on Teaching is funded primarily by the Teaching Division of the National Institute of Education, United States Department of Health, Education, and Welfare. The opinions expressed in this publication do not necessarily reflect the position, policy, or endorsement of the National Institute of Education. (Contract No. 400-76-0073)

Abstract

A taxonomy is proposed for measuring the content of instruction for fourth-grade mathematics. The results of applying the taxonomy to four standardized mathematics tests are presented.

MEASURING THE CONTENT OF INSTRUCTION[1]

William H. Schmidt[2]

For the past two years, the External Factors Group at the Institute for Research on Teaching has been studying teacher decisions regarding the content of instruction. As part of our work, we had to develop a method for describing the variety of content taught in fourth-grade mathematics. On the assumption that the items in standardized achievement tests of mathematics reflect that variety, we carried out an analysis and classification of items on the fourth-grade level Stanford Achievement Test. The result of that content analysis and classification was the development of a taxonomy (see Figure 1). The taxonomy is designed to classify the content of various curricular materials used in mathematics (including tests, workbooks, textbooks, handouts, and worksheets), as well as the content of actual classroom instruction.

The taxonomy is defined by the intersection of three factors: (1) mode of presentation (3 levels), (2) nature of the material (13 levels), and (3) operations (12 levels). The intersection of these three factors results in 468 cells.

In some respects, the taxonomy may seem overly detailed, while in

---

[1]This paper was presented at the annual meeting of the American Educational Research Association, 1978, in Toronto. A discussion on measuring instructional content is also included in an expanded report of work by the External Factors Group, "Impact on What? The Importance of Content Covered," IRT Res. Ser. No. 2.

[2]William H. Schmidt is a professor of educational psychology and senior researcher in the Institute for Research on Teaching.

others, it may appear to gloss over important distinctions. Our goal was to provide a level of detail sufficient for describing teacher decisions about the content of instruction. Hence, it is important that our taxonomy have enough detail that instruction can be directed to some cells and not to others. The taxonomy has been reviewed by several elementary school teachers, and their reviews have generally satisfied us that the taxonomy has this level of detail.

The first taxonomy factor -- Mode of Presentation -- distinguishes between items which present essential information in graphs, figures, and tables, and those which do not. For the latter items, a further distinction is made between items which specify the operation required for solution and those which do not (e.g., the typical story problem).

The second factor -- Nature of the Material -- has several levels; these levels, which are not mutually exclusive, are ordered in complexity. In using the taxonomy, we classify an item at the highest appropriate level of complexity. In ascending order of complexity, the levels are: (1) single digits, (2) single and multiple digits, (3) multiple digits, (4) single fractions, (5) multiple fractions, (6) decimals, (7) percents, (8) alternative number systems (e.g., Roman numerals, clock arithmetic), (9) place value, (10) number sentences, (11) algebraic sentences (unknown quantities not isolated by an equal sign), (12) conversion from one scale of measurement to another, and (13) geometric figures.

The third factor -- Operations -- also includes levels which are not mutually exclusive and, again, items are classified at the highest appropriate level of complexity. From the least to most complex, the levels are: (1) add, (2) subtract without borrowing, (3) subtract with borrowing, (4) add or subtract fractions without a common denominator, (5) multiply, (6) divide without remainder, (7) divide with remainder,

(8) combination (more than one of the basic arithmetic operations), (9) grouping (use of parentheses), (10) identify equivalents (e.g., select the figure with a fourth of its area shaded), (11) identify rule (e.g., number series problems), (12) identify terms (essentially vocabulary).

Classifying the Content of Standardized Tests

The popularity of standardized tests for evaluating programs and student progress makes knowledge of their content an important goal in and of itself. Toward that end, we classified the fourth-grade mathematics content of the four most widely-used standardized tests: the Stanford Achievement Test (SAT), the Iowa Test of Basic Skills (Iowa), the Metropolitan Achievement Test (MAT), and the California Test of Basic Skills (CTBS).[3]

The items in the mathematics subtests of all four standardized test batteries were independently classified by three raters. Assuming that agreement between two of the three raters made an item classifiable, 98% of all items could be classified.

Inter-rater reliabilities are reported in Table 1, by test battery, subtest, and dimension of the taxonomy. The cell entries give the

---

[3] Iowa Tests of Basic Skills (1971); Level 10; Tests M-1, M-2, and appropriate items on W-2.

Metropolitan Achievement Tests (1970); Elementary Level; Tests 5, 6, & 7.

Stanford Achievement Tests (1973), Primary Level III (3rd Grade), Intermediate Level I (4th grade), and Intermediate Level II (5th grade); Tests 4, 5, & 6.

California Tests of Basic Skills (1968); Level II; Tests 6 & 7.

percent of possible pairs of raters agreeing; for each item, agreement among all three raters counted as three of three possible pairs and agreement between two raters counted as one of three. Entries in the C columns represent agreement as to the exact cell in the matrix. As might be expected, the computation subtests were described with the greatest accuracy -- 90% or more agreement at the exact cell level. Items in the concepts subtests were most difficult to describe, with exact cell agreements at only 60%. Three of the four tests were nearly equal in terms of how accurately they could be described; the Iowa was the exception, because it did not contain a subtest devoted to computation.

The percentages of items at each level of every dimension are presented in Table 2 for each test. We classified items by reviewing the independent decisions of the three raters and resolving disagreements to their mutual satisfaction. The reliabilities reported in Table 1 represent, therefore, a strong lower bound to that for data in Table 2. In one sense, the data in Table 2 may be misleading in that the percentages reported for the marginals of the taxonomy could be in agreement, but there would still be no overlap in classification of items from the different tests at the cell level. To the extent that differences occur on the marginals, however, the tests do differ in content and at a rather low level of detail.

Three of the four tests appeared quite similar in terms of Mode of Presentation. The Iowa again was the exception; it had a substantially larger proportion of items which presented essential information in the form of graphs, figures, and tables. This

difference was due, in part, to the absence of a computation subtest on the Iowa but not entirely, since the raw number of such items was considerably greater on the Iowa, as well. Excluding the Iowa, about 20% of the items on each test involved graphs, figures, or tables, and slightly less than a third of the items required the respondent to figure out the necessary operation (for the most part, story problems).

In the area of the Nature of Materials, there were more similarities than differences among the four test batteries. Still, some important differences were found. For example, the subtotals for the three levels involving whole numbers varied from 39% to 66% with the SAT having the highest percentage.[4] The other frequently represented levels were algebraic sentences at roughly 10%, and essential units of measurement, which ranged from a low of 7% on the SAT to a high of 15% on the MAT. Percents, alternative number systems, and geometric figures were not emphasized on any of the tests.

On the Operations factor, the tests were quite similar in the percentages of items involving subtract without borrowing (6% - 8%), add or subtract fractions without a common denominator (0% - 2%), divide with remainder (1%), and combinations (6% - 8%). For the remaining levels, there were modest to strong differences. The MAT, for example, had about eight percentage points more addition items than any of the other tests. The Iowa had at least five percentage points fewer multiplication items than the other tests. Grouping was tested on the SAT and Iowa, but not at all on either the MAT or the CTBS.

To provide some sense of how the tests varied in content across grade levels, the third- and fifth-grade levels on the SAT were also

---

[4]For the SAT, a percent is about .9 of an item. An item is equivalent to approximately .2 of a grade equivalent near the middle of the norm distribution of scores on the SAT math subtests.

analyzed. The results, which are reported in Table 3, were based on resolution of any disagreements between two independent raters, both of whom were also raters for the data in Table 2. As the Table indicates, the distribution of items at the Mode of Presentation level remained nearly identical from the third-grade level to the fifth-grade level. Under Nature of Material, the percent of items classified as single digits and place value decreased with increasing grade level, while the percent of items classified as fractions, decimals, and percents increased. Surprisingly, the percent of items classified as algebraic sentences held quite constant at about 10%.

The data in Tables 2 and 3 represent descriptions of mathematics content mode using only the marginals of the taxonomy. The data in Figure 2 are item distributions across the cells of the taxonomy for the Concepts subtest of the SAT and the MAT. The figure shows that across the two subtests, items fell into 47 different cells, with only 7% to 15% of those cells common to both tests.

Although the cell-level analysis was more dramatic, sizable differences were also reflected in comparisons on the marginals. For example, 12% of the MAT items were classified Operation Not Specified, while there were no such items on the SAT. Twenty-three percent of the MAT items involved essential units of measurement compared to only 6% of the SAT items. The SAT had larger percentages of items classified as grouping (6% to 0%), identify rule (19% to 7%), and identify term (22% to 12%).

Concluding Remarks

It is clear from these content analyses that a total score on any of the subtests represents an aggregate across many content areas. These aggregates might well vary in their sensitivity to any given mathematics instruction. Furthermore, there is sufficient variance in content across tests to suggest that some may be more relevant than others to a given instructional program (i.e., a teacher should carefully consider how closely the "implied curriculum" of the test matches the math curriculum of the class in which the test is to be used.)

The results of our analyses indicate that the taxonomy is sufficiently detailed to denote rather sizable differences in terms of content covered among the four most commonly-used standardized tests. It appears, therefore, that the taxonomy holds great promise for describing the variety of content taught in fourth-grade mathematics.

Classification of _____

By _____ Date _____

MODE OF PRESENTATION

Nature of the Material / Operation	Graphics, Figures, Tables or Physical Objects												Operation(s) Specified												Operation(s) Not Specified (Story Problems)											
	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12
Whole Numbers — single digits																																				
Whole Numbers — single digit and multiple digit																																				
Whole Numbers — multiple digits																																				
Fractions — single																																				
Fractions — multiple																																				
Decimals																																				
Percents																																				
[a] Alternate Number Systems																																				
Place Value																																				
Sentences — Number																																				
Sentences — [b] Algebraic																																				
[a] Essential Units of Measurement																																				
[a] Geometric Figures																																				
Other																																				

Operations

1. Add
2. Subtract w/o borrowing
3. Subtract with borrowing
4. Add or subtract fractions
5. Multiply
6. Divide w/o remainder
7. Divide with remainder
8. Combination
9. Grouping
10. Identify Equivalents
11. Identify Rule (Order)
12. Identify Terms[a]

[a] Be sure to *identify specifics* on attached page.

Figure 1.

Classification of X - SAT + O - MAT Concepts

By _____ Date _____

MODE OF PRESENTATION

Nature of the Material / Operation	Graphics, Figures, Tables or Physical Objects												Operation(s) Specified												Operation(s) Not Specified (Story Problems)											
	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12
Whole Numbers — single digits										OO	O						OO				X	X	X	X												
Whole Numbers — single digit and multiple digit						O						X	O				OO				X		XX	X												
Whole Numbers — multiple digits										O					OO								XX O					O								
Fractions — single										X							X						X													
Fractions — multiple										OO																										
Decimals										X																										
Percents																						X														
a Alternate Number Systems										O			O									XXOO OOO	O	X												
Place Value										X																										
Number Sentences										XX O																										
bAlgebraic													X				X			O		OO		X O												
a Essential Units of Measurement	O									X O				X			O			O		O					O			O				OO		
a Geometric Figures						X						XX OOO																								
																								O												
Other																																				

Operations

1. Add
2. Subtract w/o borrowing
3. Subtract with borrowing
4. Add or subtract fractions
5. Multiply
6. Divide w/o remainder
7. Divide with remainder
8. Combination
9. Grouping
10. Identify Equivalents
11. Identify Rule (Order)
12. Identify Terms*

* Be sure to *identify specifics* on attached page.

Figure 2: Matrix Distribution of Items on Concepts Subtests of the SAT and MAT.

Inter-rater Agreement *

	Computation				Concepts				Problem Solving				Study Skills				
	D_1	D_2	D_3	C	D_1	D_2	D_3	C	I_1	D_2	D_3	C	D_1	D_2	D_3	C	cell average
SAT	100	98	93	92	94	79	83	63	97	86	93	78	100	83	83	83	79
MAT	100	92	98	90	85	88	73	64	100	82	94	78					77.3
IOWA					81	84	70	54	83	90	98	78	100	82	58	50	60.7
CTBS	100	99	99	99	91	87	89	73	100	80	93	75	100	82	71	69	79.0
cell average				93.7				63.5				77.3				67.3	

* Entries are percent of possible pairs of three raters agreeing

D_1 mode of presentation

D_2 nature of material

D_3 operations

C cell of the matrix

Table 1

Table 2

ITEM DISTRIBUTIONS FOR EACH FACTOR ACROSS TESTS*
FOURTH GRADE LEVEL

	IOWA	MAT	SAT	CTBS
I. Mode of Presentation				
- graphs, figures, tables, etc.	43	15	21	19
- operation(s) specified	29	52	53	59
- operation(s) not specified	29	32	27	22
	(N=84)	(N=115)	(N=116)	(N=113)
II. Nature of Material				
- single digits	12	15	20	2
- single and multiple digits	12	20	23	18
- multiple digits	24	19	22	19
- total -- whole numbers	47	54	66	39
- single fraction	6	4	5	7
- multiple fractions	5	3	-	7
- decimals	6	5	4	10
- percents	-	-	1	6
- alter, number systems	-	2	1	-
- place value	8	3	5	4
- number sentences	6	1	2	-
- algebraic sentences	8	10	8	12
- essen. units meas.	10	15	7	11
- geometric figures	2	3	3	2
- other	1	1	-	2
III. Operations				
- add	12	21	13	14
- subtract w/o borrowing	8	8	6	8
- subtract with borrowing	11	11	6	5
- add or subtract fractions w/o common denominator	1	-	-	2
- multiply	11	19	16	17
- divide w/o remainder	6	9	15	14
- divide with remainder	1	1	1	1
- combination	8	6	7	7
- grouping	2	-	5	-
- identify equivalents	20	18	16	15
- identify rule (order)	11	3	9	12
- identify terms	8	5	6	4

*entries are percents

Table 3

ITEM DISTRIBUTIONS FOR EACH FACTOR ACROSS GRADES *
STANFORD ACHIEVEMENT TEST

	3rd	4th	5th
I. Mode of Presentation			
- graphs, figures, tables, etc.	18	21	18
- operation(s) specified	59	53	55
- operation(s) not specified	23	27	28
	(N=96)	(N=116)	(N=120)
II. Nature of Material			
- single digits	26	20	13
- single and multiple digits	22	23	21
- multiple digits	13	22	15
- total -- whole numbers	60	66	49
- single fraction	4	5	8
- multiple fractions	-	-	7
- decimals	-	4	3
- percents	-	1	3
- alter. number systems	2	1	3
- place value	9	5	5
- number sentences	2	2	2
- algebraic sentences	10	8	10
- essen. units meas.	6	7	6
- geometric figures	2	3	4
- other	3	-	2
III. Operations			
- add	13	13	10
- subtract w/o borrowing	10	6	2
- subtract with borrowing	9	6	8
- add or subtract fractions w/o common denominator	-	-	1
- multiply	17	16	13
- divide w/o remainder	10	15	8
- divide with remainder	-	1	6
- combination	10	7	13
- grouping	1	5	8
- identify equivalents	18	16	20
- identify rule (order)	7	9	2
- identify terms	4	6	12

* entries are percents